

# Comment éviter les coûts liés au surdimensionnement d'infrastructure de centres de données et de salles réseaux

**White Paper # 37**



## Résumé de l'étude

L'infrastructure physique et électrique des centres de données et des salles réseaux est en général surdimensionnée de plus de 100 %. Ce document présente les statistiques portant sur le surdimensionnement, quantifie les coûts induits et étudie les raisons fondamentales du phénomène. Il présente et décrit une architecture et une méthode destinées à éviter le surdimensionnement.

# Introduction

Ce document démontre que le surdimensionnement est le poste de surcoût individuel le plus lourd habituellement rencontré dans l'infrastructure des centres de données et des salles réseaux. L'infrastructure physique et la puissance réellement utilisées dans un centre de données ou une salle réseaux sont en général largement inférieures à 50 %. Cette capacité inutilisée des centres de données et des salles réseaux est un coût d'investissement qu'il est possible d'éviter, de même qu'il est possible d'éviter les coûts d'exploitation et de maintenance associés.

Ce document est composé de trois parties. Il établit tout d'abord une description des faits et des statistiques liés au surdimensionnement. Il se penche ensuite sur les raisons du phénomène. Il propose enfin une architecture et une méthode visant à éviter ces coûts.

## Faits et statistiques liés au surdimensionnement

Quiconque a fréquenté l'univers des technologies de l'information ou de l'installation d'équipements a déjà pu observer dans tout centre de données une certaine proportion d'espace ou de puissance inutilisée ou d'autres éléments d'infrastructure qui restent sous-exploités. Afin de quantifier ce phénomène, il convient de définir clairement les termes que nous allons utiliser.

### Définitions liées au surdimensionnement

Pour les besoins du présent document, les termes suivants répondront aux définitions ci-dessous :

Terme	Définition
Durée de vie prévue	La durée de vie globale prévue pour le centre de données. En général 6 à 15 ans, 10 ans étant la durée typique.
Capacité de la salle	La charge maximale que la salle est capable de supporter. L'installation d'alimentation et de refroidissement nécessaire pour assurer cette capacité peut être installée en totalité ou en partie dès le départ.
Capacité installée	La capacité de charge supportée par l'installation d'alimentation et de refroidissement installée. Égale ou inférieure à la capacité de la salle.
Charge prévue	La puissance nécessaire estimée au moment de la commande du système, et sur sa durée de vie. Évolue en général avec le temps en augmentation par rapport à la charge commandée.
Charge réelle	La puissance nécessaire réelle au moment de la commande du système, et sur sa durée de vie. Évolue en général avec le temps en augmentation par rapport à la charge commandée.

## Confirmer des hypothèses

Afin de réunir et d'analyser les données liées au surdimensionnement, APC a observé les utilisateurs et élaboré un modèle simplifié pour décrire des plans de capacité d'infrastructure de centres de données. Le modèle considère ce qui suit :

- La durée de vie prévue d'un centre de données est de 10 ans.
- Un plan de centre de données comporte une capacité d'alimentation finale prévue et une Puissance nécessaire estimée au démarrage.
- Dans le cours du cycle de vie normal d'un centre de données, on prévoit que la Charge prévue augmentera de façon linéaire, en partant d'une charge de démarrage prévue pour atteindre une valeur finale ultime à mi-parcours de son cycle de vie.

Le modèle défini ci-dessus donne naissance au modèle de planification du schéma 1. Nous considérons ce modèle comme représentatif de la façon dont les systèmes sont planifiés.

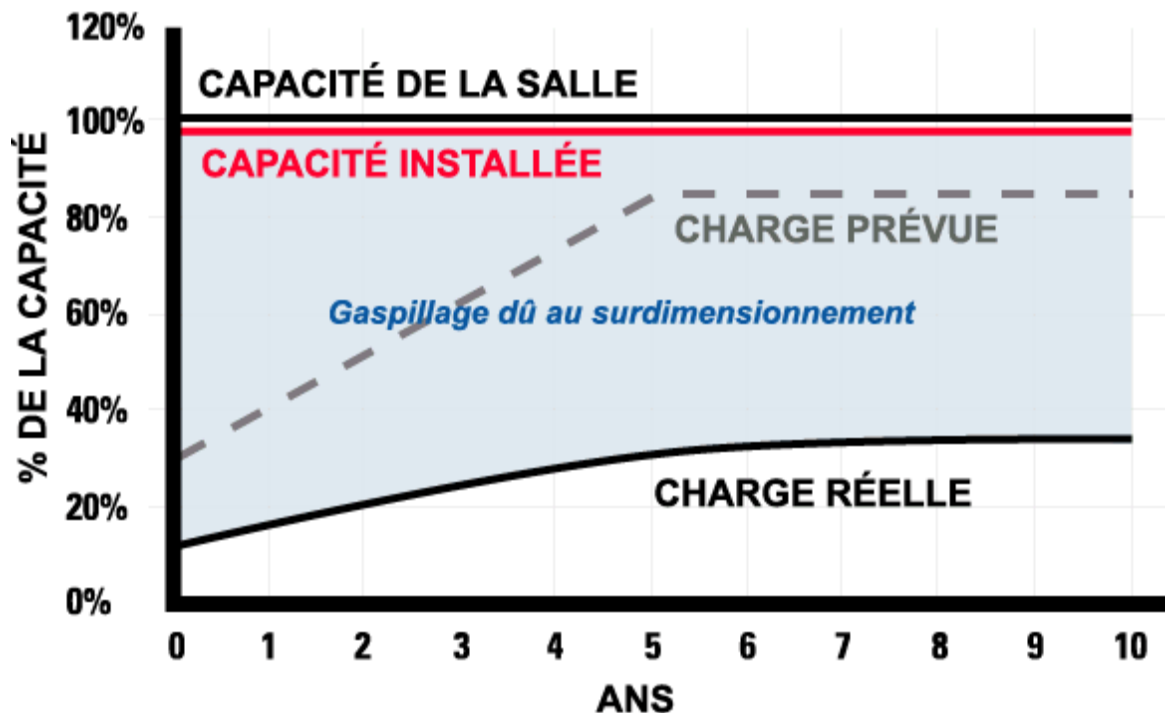


Figure 1 – Capacité et besoins d'alimentation prévus sur la durée de vie d'un centre de données

Le schéma montre un cycle typique de planification. La Capacité installée des équipements d'alimentation et de refroidissement installés est égale à la Capacité de la salle. Autrement dit, le système est construit en

totalité dès le départ. Le plan prévoit que la Charge prévue du centre de données ou de la salle réseaux débutera à 30 % et atteindra progressivement la Charge prévue finale. Cependant, la Charge réelle au démarrage se révèle en général inférieure à la Charge prévue au démarrage et n'augmente progressivement que jusqu'à une Charge réelle finale considérablement plus faible que la Capacité installée (on remarquera que la Capacité de puissance électrique nominale du matériel réellement installé pourra être supérieure à la Capacité installée en raison de l'existence de doublons ou de marges de déclassement souhaitées par l'utilisateur).

## Données provenant d'installations réelles

Pour comprendre à quel point les installations réelles étaient véritablement surdimensionnées, APC a recueilli des données auprès de nombreux utilisateurs, en effectuant une étude sur des installations actuelles et en interrogeant des utilisateurs. Les résultats ont montré que la Charge de démarrage prévue représentait en général 30 % de la Charge ultime prévue. Il est apparu d'autre part que la Charge de démarrage réelle représentait en général 30 % de la Charge de démarrage prévue et que la Charge réelle ultime représentait en général environ 30 % de la Capacité installée. On trouvera ces données résumées dans le schéma 1. Un centre de données de taille moyenne est donc finalement surdimensionné d'un facteur 3 dès sa conception. Au moment de la commande, le surdimensionnement est encore plus sensible, en général d'un facteur 10.

## Surcoûts liés au surdimensionnement

Les coûts de durée de vie liés au surdimensionnement peuvent être divisés en deux catégories : les coûts d'investissement et les coûts d'exploitation.

Les surcoûts liés à l'investissement sont représentés par les zones ombrées du schéma 1. La zone ombrée du schéma symbolise la part de la capacité du système qui reste non utilisée dans une installation de taille moyenne. La capacité excédentaire se traduit directement par un surcoût d'investissement. Ce dernier comprend le coût des équipements de puissance d'alimentation et de refroidissement, ainsi que les coûts d'investissement en conception et en installation, y compris le câblage et l'acheminement des fluides.

Dans un centre de données typique de 100 kW, les systèmes d'alimentation et de refroidissement représentent un coût d'investissement de l'ordre de 500 000 €, soit 5 € par Watt. Cette analyse montre qu'environ 70 % de cet investissement, c'est-à-dire 350 000 €, est gaspillé. Dans les premières années, le taux de gaspillage est encore plus élevé. Si l'on effectue le calcul de ce coût sur la durée, on trouve que la perte moyenne liée au surdimensionnement approche 100 % du coût total d'investissement du centre de données ! Autrement dit, les intérêts du capital d'origine suffiraient presque à eux seuls à couvrir les besoins réels d'investissement.

Les surcoûts de durée de vie liés au surdimensionnement concernent également les dépenses d'exploitation des installations. Ces coûts sont ceux des contrats de maintenance, des consommables et de l'électricité. Les coûts de maintenance sont en général légèrement inférieurs au coût d'investissement sur la durée de vie d'un centre de données ou d'une salle réseaux, lorsque la maintenance du matériel est

assurée conformément aux instructions du constructeur. Puisque le surdimensionnement entraîne la sous-utilisation d'une partie du matériel, qui doit néanmoins être entretenu, une large part des coûts de maintenance se révèle donc être inutile. Dans le cas de notre exemple d'un centre de données de 100 kW, ce surcoût est de l'ordre de 250 000 € sur la durée de vie du système.

Les surcoûts d'électricité sont loin d'être négligeables lorsque les centres de données ou les salles réseaux sont surdimensionnés. La perte provoquée par la marche au ralenti du système d'alimentation d'un centre de données ou d'une salle réseaux est de l'ordre de 5 % de la puissance nominale. Lorsque l'on y intègre les coûts de refroidissement, cette perte atteint 10 %. Pour un centre de données de 100 kW, surdimensionné de façon habituelle, le gaspillage d'électricité sur les 10 ans de vie du système est de l'ordre de 600 000 kWh, soit environ 55 000 €.

Le surcoût total, sur la durée de vie du centre de données ou de la salle réseaux, avoisinera en moyenne 70 % du coût de l'infrastructure d'alimentation et de refroidissement. Cela représente une somme qui pourrait en théorie être recouvrée si l'infrastructure pouvait s'adapter et se transformer pour répondre aux besoins réels.

Pour de nombreuses entreprises, le gaspillage d'argent investi et dépensé finit par coûter cher en opportunités perdues, une perte qui peut se révéler bien plus important que les dépenses effectuées. Des sociétés d'hébergement Internet ont ainsi fait faillite parce que le capital inutilisé bloqué sur une installation les empêchait de saisir une nouvelle opportunité qui s'offrait et de se déployer.

## Pourquoi le surdimensionnement existe ?

Les données indiquent que le surdimensionnement des infrastructures de centre de données et de salles réseaux est un phénomène à la fois de grande ampleur et d'un montant très variable. Naturellement, la question se pose de savoir si ce surdimensionnement est planifié et volontaire, s'il est dû à une mauvaise planification ou s'il tient à des raisons de fond.

### Un surdimensionnement planifié

Les rencontres avec des responsables d'installations représentatives montrent que les centres de données sont conçus pour pouvoir répondre aux besoins d'alimentation futurs maximums que la charge doit induire. La Capacité de la salle et la Capacité installée sont prévues légèrement supérieures à la Charge ultime prévue. Il est courant de déclasser le système électrique et de n'utiliser qu'une fraction de sa capacité nominale, par exemple 80 %, en pensant que faire tourner le système en dessous de sa pleine puissance permettra d'optimiser sa fiabilité globale.

La pratique qui consiste à prévoir une Capacité installée supérieure à la Charge prévue ultime d'un centre de données donne ce que l'on constate dans le schéma1. Il s'agit ici d'une forme de surdimensionnement intentionnelle et organisée. Bien que ce type de surdimensionnement soit une des formes de sous-utilisation possibles, il n'est pas pour autant le principal responsable des surcoûts globaux.

## Processus et défauts de planification

Le processus de planification habituel des centres de données et des salles réseaux s'appuie sur un certain nombre d'à priori quant aux besoins futurs, qui supposent notamment que :

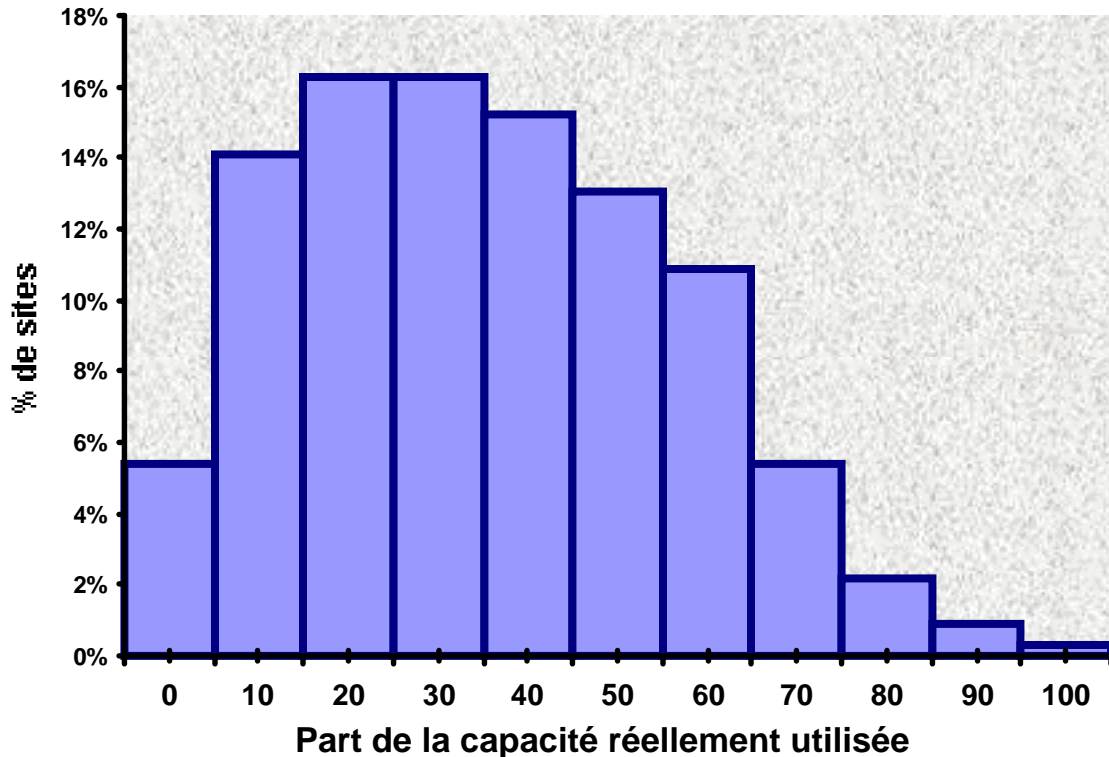
- Le coût induit par la sous-capacité d'un centre de données ou d'une salle réseaux est très élevé et doit être évité.
- Augmenter la capacité d'un centre de données ou d'une salle réseaux au milieu de sa durée de vie s'avère très coûteux.
- Les opérations liées à la mise à niveau de la capacité d'un centre de données ou d'une salle réseaux en cours de vie entraînent un risque élevé et inacceptable d'interruption de service.
- Tous les travaux de conception et de planification de la capacité ultime d'un centre de données ou d'une salle réseaux doivent être menés de façon préalable
- Il est certain que la charge du centre de données ou de la salle réseaux augmentera mais il est impossible de prédire cette augmentation avec certitude.

L'ensemble de ces à priori fait que les centres de données ou les salles réseaux sont planifiés, conçus et construits dès le départ pour répondre à des besoins inconnus et que leur capacité est prévue, à titre de précaution, en fonction du scénario représentant la fourchette la plus haute.

### Les raisons fondamentales du surdimensionnement

Le processus de planification engendre des plans qui, en moyenne, aboutissent à un niveau d'utilisation fort médiocre si l'on en juge par leurs résultats concrets. On doit donc le considérer comme un échec sur le plan économique. Pourtant, l'examen ci-dessus du processus de planification ne révèle aucun défaut fondamental. Cette contradiction apparente peut être résolue si l'on se penche de plus près sur les données et sur les contraintes pesant sur le processus. Le schéma 2 montre la distribution de la proportion de l'utilisation ultime d'installations réelles, c'est-à-dire la Charge réelle ultime divisée par la Capacité installée ultime. Une étude de ces données nous donne les indications suivantes :

- La valeur attendue de la part réellement utilisée avoisine 30 %
- La valeur attendue de la puissance électrique excédentaire ou inutile est de 70 %
- La part réellement utilisée varie considérablement, ce qui suggère que la capacité moyenne de prédire l'avenir lors du processus de conception est très faible.
- Si la Capacité installée était régulièrement fixée à la valeur attendue de 30 %, au lieu des valeurs habituellement choisies, 50 % des systèmes seraient incapables de faire face aux besoins induits par la charge au cours de leur vie.
- La technique actuelle d'évaluation de la taille est un compromis logique par lequel un système surdimensionné préserve la forte variabilité de sa Charge réelle ultime en limitant la possibilité de se retrouver dans l'incapacité d'assurer la charge requise au cours de sa vie.



*Schéma 2 – Part finalement utilisée dans un centre de données typique*

La conclusion surprenante est que, étant données les contraintes de conception et l'imprévisibilité des besoins électriques futurs, la méthode actuelle de planification des centres de données et des salles réseaux est logique. Si, pour l'entreprise, le coût de création d'un centre de données ou d'une salle réseaux incapable de faire face aux besoins est élevé, alors, dans le cadre d'une approche classique de création des centres de données et des salles réseaux, la meilleure manière de limiter le coût global attendu du système est de le surdimensionner très largement.

## Une architecture et une méthode pour éviter le surdimensionnement

L'incertitude fondamentale quant aux besoins futurs au moment du processus de planification des infrastructures de centre de données et de salles réseaux constitue un obstacle insurmontable qui ne peut être franchi qu'en étant capable de prédire l'avenir. La solution évidente consiste donc à fournir des infrastructures de centres de données et de salles réseaux susceptibles de s'adapter à une demande par essence imprévisible.



## Les obstacles à l'adaptabilité

La question qui se pose naturellement, une fois mesurée l'ampleur du problème du surdimensionnement, est la suivante : pourquoi construire à l'avance des infrastructures de centre de données et de salles réseaux, plutôt qu'au fur et à mesure pour suivre l'évolution des besoins réels ?

Le fait est que de nombreux centres de données sont effectivement conçus en intégrant une certaine part de croissance. Par exemple, le déploiement de baies de matériel s'effectue fréquemment par phases successives, de même que celui de la tranche finale de l'alimentation électrique des centres de données. Dans certains cas, le déploiement d'un module onduleur redondant peut se faire par étapes. De telles approches permettent de réaliser des économies sur la durée totale de vie d'un centre de données. Néanmoins, dans de nombreux cas, les surcoûts liés à l'installation de ce matériel dans un deuxième temps sont largement supérieurs à ce qu'ils auraient été si l'équipement avait été installé dès le départ, si bien que de nombreux planificateurs préfèrent recourir à une installation préalable complète. C'est pourquoi on ne voit dans la pratique se réaliser qu'un très faible volume des économies potentiellement envisageables.

## Méthode et approche pour créer une infrastructure évolutive

La situation idéale serait de pouvoir fournir une méthode et une architecture susceptibles de s'adapter en permanence à l'évolution des besoins. Celles-ci présenteraient les caractéristiques suivantes :

- Le travail initial de conception du centre de données et de la salle réseaux serait fortement réduit, voire supprimé
- L'infrastructure électrique du centre de données ou de la salle réseaux serait fournie sous forme de modules évolutifs préfabriqués
- Les composants pourraient être véhiculés en empruntant des couloirs ordinaires et des ascenseurs publics et pourraient se brancher sans opérations de câblage sur circuits actifs
- Les travaux spécifiques de préparation du site, par exemple la surélévation du sol, n'auraient plus lieu d'être
- Le système serait capable de fonctionner dans une configuration N, N+1 ou 2N sans modification
- Les travaux d'installation tels que câblage, percement, découpe deviendraient inutiles
- Les procédures spéciales d'autorisation ou réglementaires ne seraient plus nécessaires pour augmenter la capacité
- Le coût d'équipement du système d'alimentation modulaire serait identique ou inférieur à celui d'un système centralisé classique
- Le coût de maintenance du système d'alimentation modulaire serait identique ou inférieur à celui d'un système centralisé classique.

## Niveaux d'adaptabilité pratiques et réalisables

Lorsqu'un système d'infrastructure physique évolutif est déployé, le gaspillage lié au surdimensionnement, tel que le montrait la zone ombrée du schéma 1, peut se trouver considérablement diminué. Le schéma 3

ci-dessous montre les économies réalisées. On remarquera que la Capacité installée n'est pas fonction de la Capacité de la salle au démarrage et qu'elle évolue pour suivre la progression de la Charge réelle.

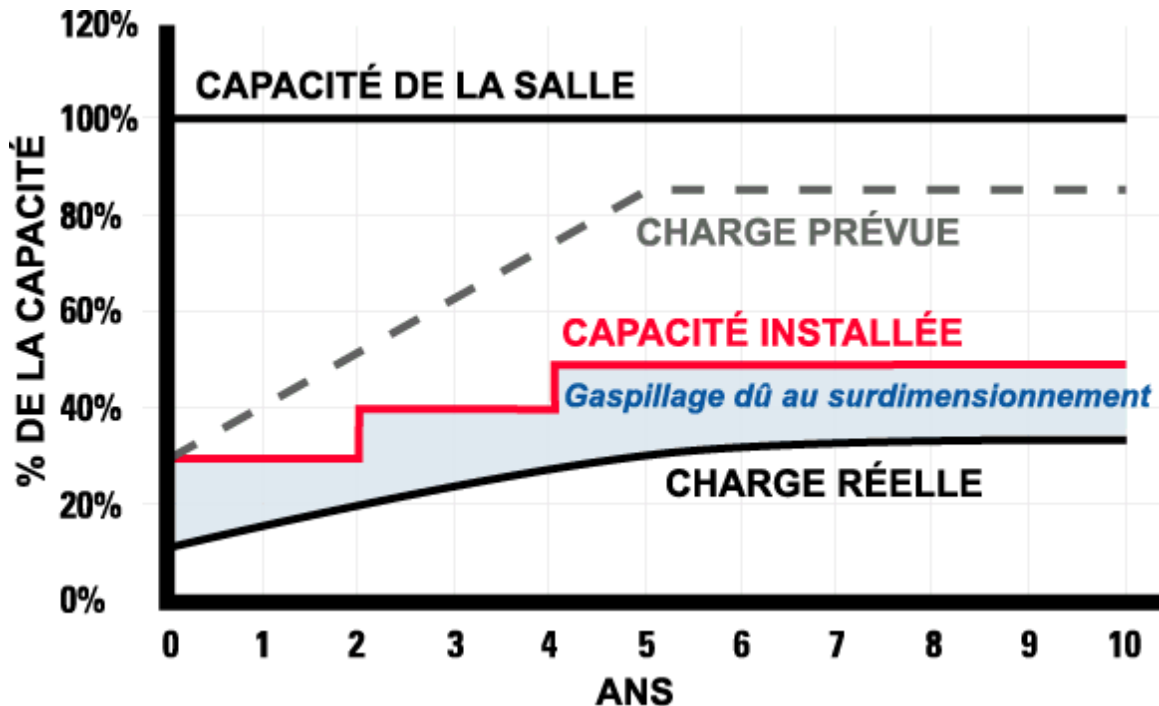


Schéma 3 – Puissance prévue et besoins sur la durée de vie d'un centre de données

Un système évolutif, capable de répondre aux besoins exprimés ci-dessus, est possible : c'est notamment le cas de l'architecture APC InfraStruXure. On ne trouvera pas ici de description complète de ce système. Dans l'architecture InfraStruXure, plus de 70 % de la capacité électrique du système peut être déployée de façon à accompagner la croissance des besoins du centre de données ou de la salle réseaux. Dans la pratique, la seule partie du système d'alimentation qui soit totalement déployée en amont est l'interrupteur principal et les armoires de distribution, dont la taille est adaptée à la Capacité ultime de la salle. L'onduleur, le système de batterie, les Unités de distribution électrique, le commutateur de bypass et le câblage de l'armoire de distribution électrique sont tous déployés de façon modulaire pour pouvoir s'adapter à l'évolution de la charge.

On notera que cette étude s'est limitée aux caractéristiques liées aux systèmes d'alimentation et de refroidissement, qui représentent les principaux postes de coûts pour les infrastructures de centres de données et de salles réseaux. Pour être exhaustive, la même analyse pourra et devra être étendue aux besoins en espace physique, en matériel de protection incendie et en sécurité.

# Conclusions

Les centres de données et les salles réseaux sont en règle générale surdimensionnés d'un facteur trois par rapport aux besoins réels. Ce surdimensionnement entraîne des surcoûts d'investissement et de maintenance, qui représentent une part non négligeable du coût global du système au cours de sa vie. L'essentiel de ces surcoûts peut être recouvré par la mise en place d'une méthode et d'une architecture capable de s'adapter de façon économique à l'évolution des besoins tout en assurant un degré élevé de disponibilité.

# Références

Mitchell-Jackson, J.D., Koomey, J.G., Nordman, B., Blazek, M., "Data Center Power Requirements: Measurements From Silicon Valley", 16 mai 2001. Thèse de Master, Energy and Resources Group, Université de Californie. Berkeley, Californie. Disponible à l'adresse <http://enduse.lbl.gov/Projects/InfoTech.htm>